

# Bilingual Education and Identity

ON-GOING RESEARCH  
VERY PRELIMINARY DRAFT

Antonio Di Paolo  
(AQR-IREA, Universitat de Barcelona)

Ramon Caminal  
(Institute of Economic Analysis, CSIC, and BSC)

Ada Ferrer-i-Carbonell  
(Institute of Economic Analysis, CSIC, and BSC)

February 2023

## Abstract

We examine the impact of a language-in education reform that introduced bilingualism in the Spanish region of Catalonia on identity formation and political preferences. We use repeated cross-section data for several waves. Multiple waves allow us to better separate the age and cohort effects. We consider different aspects of individual identity: linguistic identity (the language that individuals consider their own) as well as national (feeling more or less Spanish than Catalan), and alternative measures of political preferences. The preliminary results indicate that the exposure to the reform might have contributed to increase the probability of adopting Catalan as the language of self-identification, especially among native Spanish speakers. However, the effect on Catalan national identity is generally negative. If anything, the reform has moderated extreme feelings and promoted the dual identity (as Spanish as Catalan). Indeed, the two dimensions of identity (linguistic and national) are only moderately correlated. Hence, the language is no longer the distinct marker of ethnic or national identity. The effect of the reform on political preferences is more ambiguous, as the sign of the coefficient varies by indicator. Surprisingly, the strongest negative effect of exposure on their support for independence is among native Catalan speakers.

*Key words:* education, identity, language, political preferences

*JEL Classification numbers:* I28, J15, Z13

## 1. Introduction

In many multilingual societies, language-in-education policies are highly controversial. Firstly, they can largely determine individual language repertoires, thus affecting economic and social outcomes. Secondly, there is a broad concern that they may also affect identity formation and even spur inter-ethnic conflict. Indeed, some of these policies are associated with “nation-building” efforts. In this paper, we examine the impact of bilingual education on identity formation in Catalonia, a region of Spain whose secession movement has recently been under the spotlight.

An essential feature of the construction of nation states in the XIX and XX centuries was the homogenization of a diverse population; in particular, the universal adoption of the national identity. This was primarily achieved by introducing state-controlled mass education, enforcing the exclusive use of the “national language” in school, implementing compulsory military service, and building infrastructures (See Alesina et al., 2021, and the historical references they cite).<sup>1</sup>

France provides a prominent example. Focusing on the language dimension, at the time of the Revolution French was only spoken by a small fraction of the population, between 10% (Hobsbawm, 1990) and 40% (Calvet, 2002). In 1860 this fraction was already higher: about 50% (Weber, 1979). In the 1880’s schooling became free and French was made the only language of instruction. By that time, school attendance ratios were very high, and French soon became the dominant language at the expense of the numerous regional languages. In some other countries, homogenization took longer or remained incomplete.<sup>2</sup>

Education policies were crucial for the adoption of a common language and a common identity. The link between language and identity was believed to be so tight that individuals were induced or forced to give up their original language and become exclusive speakers of the national language. Thus, nation building implied reduced language diversity (“One nation, one language”).

The expansion of international trade and migration induced individuals to acquire second languages, but such bilingual groups rarely threatened the predominance of the national language. In particular, second-generation immigrants in the US tended to assimilate to the domestic language and culture (Alba et al., 2002). However, there were exceptions, sometimes involving education policies. An interesting early example concerns the German-speaking community in the Midwestern US at the beginning of the

---

<sup>1</sup> The incentives for investing in nation building were both political: responding to internal and external threats (Alesina et al., 2020), and economic: the creation of large, unified domestic markets (Alesina et al., 2000).

<sup>2</sup> At the unification of Italy in 1860, the Italian language was spoken at most by 10% of the population (Duggan, 2007). In this case, school attendance and literacy rates remained low for a number of decades. The case of Spain is briefly discussed below.

XX century. Before 1917 bilingual education was common in many states with a large fraction of German immigrants. After 1917, when the US joined World War I, several states banned German as a language of instruction.<sup>3</sup> Fouka (2020) documents how such a policy shift, rather than resulting in faster assimilation, “contributed to a cultural backlash and greater isolation of the German community from mainstream American culture”. The strengthened influence of school on identity was compensated by the extra effort of the parents in transmitting their ethnic identity to their children.

Nowadays, bilingual education (English and Spanish) is still a contentious policy issue in the US. In fact, its legal status has changed several times in California, the state with the largest Spanish speaking community. Supporters of the “English only” policy fear that bilingual education will delay the assimilation of the Spanish speaking population. Similar reactions can also be observed in other countries, not only with respect to the languages of the migrants, but also to domestic minority languages. In other words, the promotion of language diversity is often perceived as an attempt to break up the homogeneity of the nation and challenge the single identity (Ricento, 2013).

In a few corners of the western world, minority groups are trying to revitalize their language and culture. This paper focuses on the case of Catalonia. Unlike other neighboring countries, Spain has not been entirely successful in enforcing a single identity.<sup>4</sup> In fact, in the XIX century Spanish authorities did not invest enough in mass education, as the interests of the country elite and the regional elites diverged (Hauk and Ortega, 2019). Whereas primary enrollment ratio in France in the 1880’s was already 75%, in Spain was only 42% (Nuhoglu Soysal and Strang, 1989).<sup>5</sup> Such limited efforts are probably behind the relative vitality of the main Spanish regional languages (Galizian, Euskera, and Catalan).

The Catalan case is particularly interesting from a research point of view for at least two reasons. First, Catalan is one of the most widely spoken languages in Europe that is not official at the country level (except in the microstate of Andorra). Second, in the early 1980’s the regional government undertook a large-scale education reform after the approval of the Language Normalization Act, LNA (*Llei de Normalització Lingüística*) of 1983, right after the end of Franco’s regime, when Catalan was forbidden. The application of such legislation to education changed the school from a monolingual system, in which Spanish was the only language, to a bilingual one, in which Catalan became the main language of instruction. Such a reform can be considered a quasi-experiment and hence it can provide the foundation for a solid identification strategy. It is precisely this education reform that will be at the center of our analysis. In the early 1980’s the Catalan society was asymmetrically bilingual. All native Catalan speakers were also fully

---

<sup>3</sup> These events coincided with the *Americanization campaign, 1914-1924*, which is considered the period of the emergence of the American national identity (Ricento, 2003).

<sup>4</sup> Cáceres et al. (2021) and Bagues and Roth (2022) document the role of compulsory military service in promoting the Spanish national identity vis-à-vis regional identities.

<sup>5</sup> See also Bandiera et al. (2018).

competent in Spanish, but a large fraction of native Spanish speakers was monolingual or only passively bilingual. The reform aimed at producing fully bilingual individuals at the end of compulsory education, regardless of their linguistic background.

Thus, one of the expected effects of the reform is the expansion of language skills, especially for students coming from Spanish speaking families. Two empirical questions naturally arise. First, did the reform actually improve language skills? Second, have these additional skills had any significant effect on economic and social outcomes? It is important to notice that the improvement of Catalan skills would be redundant from a communicative point of view, since everyone was fluent in Spanish. Hence, any effect must be attributed to the non-communicative aspects of languages: emotional attachment, culture, identity.

Our previous research has provided positive answers to both questions. In Caminal and Di Paolo (2019), and Caminal, Cappellari, and Di Paolo (2021) we show that the education reform did improve the oral and written Catalan skills of native Spanish speakers, and the written skills of native Catalan speakers, without significantly affecting the Spanish skills of either community. Moreover, these additional skills caused an increase in earnings (especially, for individuals with non-Catalan origins (Cappellari and Di Paolo 2018), raised the frequency of mixed couples (Caminal and Di Paolo, 2019) and fostered the intergenerational transmission of Catalan (Caminal et al., 2021).

The strong link between language and identity, discussed above, suggests that the reform may have also altered individual identities. In particular, the change in the language of instruction, together with other aspects of the reform, like new textbooks or changes in the composition of the teaching body, might have prompted the adoption of the Catalan identity among native Spanish speakers. In addition, it might have intensified the Catalan identity of the native Catalan speakers. In other words, the main research question is whether bilingual education can be an effective instrument of “nation-building”, or at least, whether it could contribute to redefine the borders of linguistic/ethnic groups.<sup>6</sup>

Naturally, this is an empirical question. Nevertheless, it is important to emphasize that the conditions under which such an education reform took place are quite different from those existing during the construction of the western nation states. We need to distinguish between intention and capacity to achieve a certain goal. Regarding intention, the existence of political and social groups whose aim is the promotion of the Catalan identity is quite obvious. Indeed, political parties in Catalonia have always competed over two dimensions: right-left, and power distribution in a two-tier government. The second dimension is about the organization of the state (centralization, decentralization, sovereignty) which is closely linked to the alternative cultural/linguistic identities (Catalan versus Spanish). Even though the electoral support of the Catalan-oriented

---

<sup>6</sup> Research on the formation of “regional” identity is scarce. An important exception is Dehdari and Gehring (2022). They show how negative experiences with nation-states, like war, occupation, and repression, affected the formation of a specific identity in Alsace-Lorraine.

parties has never been significantly above 50%, the largest Catalan-oriented party (CiU) controlled the regional government over the period 1980-2003.

Regarding capacity, a few observations may be useful. First, the support for the education reform has always been very broad in the Catalan Parliament. The LNA of 1983 was approved by almost unanimity (one abstention, zero negative votes). Similarly, another important piece of language legislation, the Law of Linguistic Policy (*Llei de Política Lingüística*) was passed in 1998 also by a wide majority.<sup>7</sup> Second, in the 1980's about one half of the Catalan residents were native Spanish speakers. Thus, they cannot be characterized as a minority.<sup>8</sup> The implementation of the law required the cooperation of teachers and was subject to the supervision of the families. Native Spanish speakers abounded in both groups. Third, Catalonia is part of the Spanish political and judicial system. In 1994, the Spanish Constitutional Court validated the language-in-education policies of the Catalan government. Fourth, the influence of the school on identity can be counterbalanced by the reaction of the families, as demonstrated by Fouka (2020).

An important predecessor of our paper is Clots-Figueras and Masella (2013), CF-M hereafter<sup>9</sup>. They use data from a cross-sectional survey of *Centro de Investigaciones Sociológicas*, conducted in 2001, and found a positive effect of exposure to the language-in-education reform during compulsory education on Catalan identity. Their identity variables are generated from the response to a question about the feeling of belonging ("Do you feel: (1) Only Spanish, (2) More Spanish than Catalan (3) Catalan and Spanish, equally, (4) More Catalan than Spanish, (5) Only Catalan"). They found a negative trend across cohorts, with young cohorts feeling less Catalan than older ones, but they claim that exposure to the LNA reform delayed such decline. They also provide evidence suggesting that voting behavior was affected by the exposure to bilingual education, fostering the support for Catalan-oriented parties. Their results on political preferences (independence of Catalonia, territorial organization of the state) were not statistically significant.<sup>10</sup>

In this paper, we reexamine the possible effects of the education reform on Catalan identity. First, we exploit multiple waves of a survey carried out by the Center of Opinion

---

<sup>7</sup> After fifteen years of practical application of the 1983 law, only two small parties voted against the 1998 law, and they did it for opposite reasons. One of them (PP, a Spanish-oriented party, with 17 seats) found the law discriminatory against the Spanish language, and the other (ERC, a Catalan-oriented party, 13 seats) believed it was too shy to rescue the Catalan language from its extinction path. The total number of seats in the Catalan Parliament was 135. Thus, the law was approved by more than two thirds.

<sup>8</sup> In the case of Indonesia in the 1980's, Bazzi et al (2019) showed that fractionalization favors the adoption of the (neutral) national language, whereas the presence of a small number of large language groups generates the opposite effect, as ethnic attachment intensifies.

<sup>9</sup> See also Aspachs et al. (2008).

<sup>10</sup> Some authors (See, for instance, Ginsburgh and Weber, 2020) cite this paper as evidence that the origins of the current conflict over Catalan secession can be traced back to the educational reform. Clearly, alternative explanations are also plausible. In particular, Casas et al. (2021) argues that judicial decisions (in particular, the ruling of the Spanish Constitutional Court of 2010) decreased the Spanish national identity and increased support for Catalan secession among Catalan citizens.

Studies of the Catalan Government to investigate political and social issues, covering the period 2004-2020. The use of multiple waves allow us to disentangle the age and cohort effects. More specifically, the effect of exposure to the language-in-education reform during compulsory education is identified by exploiting variation by birth cohort. By using repeated cross-sections, we observe individuals born in the same year with different ages when answering the survey, which is crucial to identify the effect of the education reform.

Second, we examine alternative definitions of identity. Questions about the feeling of belonging, like the one used by CF-M, are informative indeed, but raise the issue of how to deal with individuals with “dual” identity (option (3) in the above list of possible answers). For instance, a shift from (2) to (3) can be interpreted as an advance of the Catalan identity, but also as a reduction in the degree of polarization, which should contribute to curtail the probability of conflict, as more individuals adopt a neutral position. Thus, on top of these standard identity definitions, whenever possible, we also exploit the concept of linguistic identity.<sup>11</sup> If a language-in-education reform, like the one explored in this paper, affected how an individual characterizes itself in relation to the surrounding social groups, then one might reasonably expect that the first dimension to be affected would be the linguistic. Indeed, one of the main themes of this paper is the degree of substitutability among alternative definitions of individual identity and how they relate to various outcomes.<sup>12</sup>

Third, we analyze the existence of heterogeneous effects of the LNA reform on identity formation along several dimensions. More specifically, we provide evidence regarding the differential effect of the language-in-education policy on the different measures of identity, not only regarding the parents’ origin, but also according to respondents’ native language.

## 2. Historical background

The two most important languages in Catalonia are Spanish (Castilian) and Catalan, both Romance languages. Historically, their relative roles in society have been very sensitive to political and institutional changes. At the beginning of the XVIII century, explicit policies aiming at the homogenization of the Spanish territories were implemented, especially in linguistic and cultural matters. As a result, the Catalan language was gradually excluded from public life. In a similar spirit, in the mid XIX century the Spanish Parliament passed an ambitious law (Moyano Law, 1857) whose goal was the creation of a compulsory, public school system with Spanish as the only language. In practice, the

---

<sup>11</sup> In our dataset some respondents claim both Spanish and Catalan as their language of self-identification. However, this group is rather small.

<sup>12</sup> The literature on ethnic divisions and development, initiated by Easterly and Levine (1997), take language as the marker of ethnicity. Unfortunately, we do not know much about the extent to which language is representative of ethnic identity.

enrollment ratio stayed low for a long time. As a result, Catalan was still widely used in Catalonia.

In the second half of the XIX century, a cultural and linguistic movement emerged (called *La Renaixença*), whose goal was the promotion of the use of Catalan in literature and, more generally, in any form of cultural expression. Soon, the movement adopted a more political attitude by supporting the use of Catalan in education and other social activities. The first manifestation of modern, Catalan-oriented political activity was the creation, at the beginning of the XX century, of *La Lliga Regionalista*, a right-wing political party, with very modest goals in terms of political autonomy for Catalonia, but that represented a clear expression of the Catalan identity, mixed with conservative values.

During the Second Spanish Republic (1931-1936) a new party (*Esquerra Republicana de Catalunya*, ERC) took over the representation of the Catalan identity. This time in combination with republican and progressive values. ERC obtained a landslide victory in the 1932 elections to the Catalan parliament,<sup>13</sup> and took over the regional government in coalition with small socialist parties. Catalan language became official in the region and was also introduced as language of instruction in the public school system. The regional government was suspended in 1934, and only restored a few months before the civil war (1936-1939).

Franco's victory implied a major setback for the Catalan language. Under his dictatorial regime (1939-1975), Spanish was the only official language and the only language used in education. Any infringement was heavily punished. Thus, Catalan was banned from public life and its use restricted to the private sphere. Nevertheless, the Catalan language was transmitted (mostly orally) across generations by many families of Catalan origin.

A strong sense of Catalan identity survived throughout these years, sustained mostly by civic groups as well as clandestine political parties and labor unions. For instance, the main publication of the largest political party of the anti-Franco movement (*Partit Socialista Unificat de Catalunya*, the Catalan communist party), called *Treball*, was written entirely in Catalan. Similarly, the main underground labor union in Catalonia was named *Comissió Obrera Nacional de Catalunya*, implicitly claiming Catalonia as a nation.

The use of Spanish in Catalonia was strongly reinforced by massive migration flows from other Spanish regions, especially in the 1960's and early 1970's. At the beginning of the 1980's Catalan was the native language of about one half of the population, who obviously were also fully competent in Spanish. In contrast, most of the native Spanish

---

<sup>13</sup> ERC obtained 52% of the vote and 56 seats out of 85.

speakers (40% of the population were born outside the region) were monolingual or only passively bilingual (Siguan, 1991).

The Spanish constitution of 1978 designed a two-tier government, in which the central government share some of the political power with regional governments. Right after the constitution of the Catalan government, the regional parliament passed in 1983 *Llei de Normalització Lingüística* (Language Normalization Act, LNA). The LNA was modified in 1998 (*Llei de Política Lingüística*, Law of Language Policy). The aim of the LNA was to make all pupils fully competent in both languages at the end of compulsory education. It also defined an integrative education model, in which children were not separated on the basis of the language spoken at home. None of these goals was affected by the 1998 reform. Thus, education experienced a gradual transition from a system from which Catalan was excluded to one in which Catalan has become the main language of instruction.<sup>14</sup>

The regional government also sought to promote the knowledge and use of Catalan using a variety of means included in the text of the LNA reform (and in some cases reinforced by the subsequent law of 1998), including Catalan-only TV channel (*Televisió de Catalunya*, TV3) and radio station, several campaigns, and language proficiency requirements for public sector jobs.

Besides the right-left dimension, Catalan politics was centered on the territorial organization of the state, the distribution of the public budget across regions, as well as on linguistic and cultural matters. During the first two decades, the linguistic policies of the regional government enjoyed a broad support in Catalonia (in sharp contrast with the reactions in the rest of Spain), and was only challenged by small, very active groups, with little impact on local parliamentary politics. In the XXI century, after the creation of a new political party, *Ciutadans*, *Partido de la Ciudadania*, language policy became much more controversial inside Catalonia, including its parliament.

Dissatisfaction with the financing of the regional government and concern about the fragile legal status of language policies led a bunch of political parties (including Catalan-oriented parties and the Catalan branch of the Spanish Socialist Party) to write a new charter for the region (*Estatut d'Autonomia*). The new draft was approved by the Catalan Parliament in 2005, but was significantly modified by the Spanish Parliament. Nevertheless, it was ratified in a referendum in 2006. In an unprecedented decision, the Spanish Constitutional Court reviewed the approved (and sanctioned in a referendum) draft and in June 2010 declared unconstitutional a number of articles and reinterpreted

---

<sup>14</sup> At least in compulsory education. Spanish was still dominant in post-compulsory secondary education. See Cappellari and Di Paolo (2018) for more details about the introduction of Catalan at school with the LNA reform.



several others. In July 2010, about one million people demonstrated in Barcelona against such a decision. Casas et al. (2021) showed that the ruling of the constitutional court was a turning point for the support for independence in Catalonia. From 2012 to 2017 massive demonstrations took place every year on September 11<sup>th</sup> (the Catalan national day). Support for independence increased to almost 50% from relatively low historical levels, 15-20% (*Centre d'Estudis d'Opinió*). In October 2017, the regional government organized an illegal independence referendum. The main leaders ended up in jail or exile.

### 3 Data and Descriptive Statistics

The empirical analysis reported in this preliminary draft is carried out using data from the survey Barometer of Political Opinion (*Baròmetre d'Opinió Política*, BOP) carried out by the Catalan Public Opinion Center (*Centre d'Estudis d'Opinió*, CEO) of the Generalitat de Catalunya. The Barometer is a periodic survey that started in 2004 and contains random samples of the population of Catalonia aged 18 or more collected using a multistage stratified procedure. Primary sample units are municipalities, randomly selected within province and city size, and the secondary sample units consists of census tracts of selected municipalities. Individuals within the secondary sample units are selected by applying to crossed quotas defined on the basis of gender, age groups, and place of birth defined according to official figures from the Population Census (*Padró Municipal d'Habitants*) of the year previous to each survey. Table 1A of the appendix contains the information about the available waves for each year (2004-2020<sup>15</sup>) and total sample size. Data were collected using the Computer Assisted Telephone Interview (CATI) method for the period 2004-2013 and the Computer Assisted Personal Interview (CAPI) method for the period 2014-2020.

The BOP survey contains several questions that reflect the respondents' national and linguistic identity. For all years and waves, it provides information about a) the feeling of belonging, coded in an ordinal scale defined as follows: (1) "Only Spanish", (2) "More Spanish than Catalan", (3) "as Spanish as Catalan", (4) "More Catalan than Spanish", or (5) "Only Catalan". Since the first wave of 2012, the survey includes a question concerning b) the self-identification language, coded as "Catalan", "Spanish", "Catalan and Spanish" or "others".<sup>16</sup>

From question (a) (feeling of belonging), we construct four different indicators of national identity: (i) the entire 1-5 scale, (ii) "Catalan only": takes value 1 if "Catalan only" (option 5) is selected, and zero otherwise, (iii) More Catalan or only Catalan": takes value one if

---

<sup>15</sup> Because the survey was not carried out in 2021 due to the COVID19 pandemic, we excluded data from the three waves of 2022 in order to avoid discontinuities in data availability.

<sup>16</sup> More specifically, the "others" category includes different combination of languages, but also the "Aranès" language, a variety of Occitan, which is spoken in the Aran Valley, a territory of Catalonia located in the North of the Pyrenees, and more accessible from Occitania than from Catalonia. Due to the very limited number of individuals reporting other combination of languages or "Aranès" as self-identification languages, we discarded these observations for the empirical analysis.

options 4 or 5 are selected, and 0 otherwise, and (iv) “Equal, More Catalan and only Catalan”: takes value 1 if options 5, 4, or 3 are selected, and 0 otherwise.

It also includes two questions regarding political preferences: c) the self-assessment about the level of autonomy of Catalonia with respect to the Central State, coded as “excessive autonomy”, “right autonomy” and “deficient autonomy” and d) the type of relationship that Catalonia should have with respect to the central state, coded as “a Spanish Region”, “a Spanish Autonomous Community”, “a Federal State within Spain” and “an Independent State”. The variable “Catalonia, independent state” takes value 1 if the respondent chooses “Independent state”, and zero otherwise. Moreover, since 2014, the BOP questionnaire includes a binary question regarding e) whether the respondent supports independence.<sup>17</sup>

Finally, besides these variables, which represent the main outcomes that we investigate in this paper, the survey also contains several sociodemographic characteristics such as gender, age, place of birth (Spanish regions for those born in the rest of Spain and country for foreign born individuals), place of birth of both parents<sup>18</sup>, province of residence (except for 2014) and completed education. From 2012, we also have information about the mother tongue of the respondent, classified in the same way than self-identification language. Moreover, there is also a question about the TV channel individuals usually watch to follow the news regarding political issues (available for all years and waves), which represents an additional piece of information that we exploit in the empirical analysis.

The main estimation sample consists of individuals born in Catalonia, and those born in other Spanish regions who migrated at age 6 or earlier. In this way, we ensure retaining only Spaniards who were entirely schooled in Catalonia, either before or after the implementation of the LNA reform. Moreover, we exclude individuals born before 1920 or aged more than 84, with the purpose of having the same age range for all the waves of the BOP survey. After cleaning for missing values in the main outcome of interest (a), which represents the one for which there is the lowest number of missing values, we end up with a pooled sample of 77678 observations.<sup>19</sup> We also retain individuals born in other Spanish regions and foreigners who migrated to Catalonia after having completed compulsory education (i.e. from age 16, considering the current duration of compulsory education), for the purpose of falsification analysis.

---

<sup>17</sup> It seems worth mentioning that the questionnaire of 2012 and 2013 includes a related question concerning voting preferences in an hypothetical referendum for the independence of Catalonia, which is not considered here because it only appears in few waves of the BOP survey and is not strictly comparable to the information available for subsequent years.

<sup>18</sup> Detailed information about the place of residence of fathers and mothers is only available since 2007. For previous waves, we only know if they were born in Catalonia, in the rest of Spain or elsewhere.

<sup>19</sup> Descriptive statistics and regressions for other outcomes and regarding control variables are carried out using the available observations only.

Descriptive statistics for the main estimation sample are reported in Table 1, for the pooled dataset and by year (i.e. aggregating all waves within each year). Descriptive statistics for the placebo sample are shown in Table A2 of the Appendix. Regarding the outcomes, we generally convert the variables of interest into dummies, considering the category that reflects the “strongest” level of attachment to the Catalan identity. The only exception is that we also retain the whole 1-5 scale of feeling of belonging. As it can be appreciated, the average of the latter variable is close to 4 (the value corresponding to “more Catalan than Spanish”) and quite stable over the years, indicating that individuals in our sample are more prone to be associated with the Catalan identity than with the Spanish one. Indeed, around 30% of the pooled sample declare that they feel only Catalan, with a moderately increasing trend over the years. As a whole, 38% of individuals in the sample indicate that Catalonia should be an independent State. This variable shows a clear increase over time, especially since 2012, which becomes more moderate for the last year. Moreover, 74% of the whole sample also consider that Catalonia has an insufficient degree of autonomy with respect to the Central State. This share is generally stable over time except for the first three years, in which is it slightly lower. The fraction of individuals who support independence in a binary choice fluctuates (over the available years) between 73% and 80%. Finally, 65% of individuals belonging to the pooled sample indicate that Catalan is their only self-identification language. However, this share declines sharply from 2012 (75%) until 2020 (55%).

**Table 1: descriptive statistics**

Variable	pooled		2004		2005		2006		2007		2008		2009		2010		2011	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
feeling of belonging (1-5)	3.835	0.942	3.662	0.888	3.708	0.913	3.694	0.912	3.797	0.899	3.749	0.895	3.776	0.909	3.742	0.944	3.799	0.940
feeling Catalan only (0-1)	0.292	0.454	0.197	0.398	0.213	0.409	0.209	0.407	0.249	0.432	0.236	0.425	0.253	0.435	0.247	0.432	0.274	0.446
feeling more Catalan than Spanish or Catalan only	0.608	0.488	0.534	0.499	0.572	0.495	0.561	0.496	0.607	0.489	0.57	0.495	0.58	0.494	0.572	0.495	0.591	0.492
feeling equal Catalan than Spanish, more Catalan or Catalan only	0.954	0.209	0.948	0.222	0.946	0.227	0.947	0.223	0.958	0.201	0.957	0.202	0.959	0.198	0.947	0.225	0.956	0.205
Catalonia independent State (0-1)	0.383	0.486	0.201	0.401	0.205	0.404	0.210	0.407	0.243	0.429	0.252	0.434	0.269	0.444	0.300	0.458	0.334	0.472
deficient autonomy of Catalonia (0-1)	0.740	0.439	0.539	0.499	0.682	0.466	0.643	0.479	0.731	0.443	0.735	0.441	0.731	0.444	0.722	0.448	0.760	0.427
binary independence preference (0-1)	0.613	0.487	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Catalan as self-identification language (0-1)	0.648	0.478	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
years of compulsory education in Catalan (0-10)	3.288	4.172	2.600	3.689	2.396	3.620	2.597	3.753	2.665	3.873	2.908	4.046	2.950	4.091	3.042	4.063	3.086	4.158
both parents born in the rest of Spain (0-1)	0.249	0.433	0.379	0.485	0.247	0.431	0.237	0.425	0.245	0.430	0.243	0.429	0.231	0.421	0.247	0.431	0.238	0.426
native Spanish speaker (0-1)	0.374	0.484	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
female (0-1)	0.507	0.500	0.511	0.500	0.500	0.500	0.503	0.500	0.510	0.500	0.506	0.500	0.509	0.500	0.507	0.500	0.512	0.500
age (18-84)	46.36	17.17	42.94	17.49	44.54	17.23	45.13	17.60	45.95	17.63	45.53	17.60	46.05	17.59	46.26	16.91	46.79	17.18
TV3 for political information (0-1)	0.518	0.500	0.604	0.489	0.645	0.479	0.586	0.493	0.547	0.498	0.534	0.499	0.548	0.498	0.537	0.499	0.543	0.498

Variable	2012		2013		2014		2015		2016		2017		2018		2019		2020	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
feeling of belonging (1-5)	3.938	0.895	4.033	0.926	4.014	0.966	3.976	0.979	3.912	0.967	3.850	1.011	3.962	0.979	3.874	0.985	3.805	0.996
feeling Catalan only (0-1)	0.319	0.466	0.389	0.488	0.389	0.488	0.385	0.487	0.346	0.476	0.332	0.471	0.375	0.484	0.340	0.474	0.311	0.463
feeling more Catalan than Spanish or Catalan only	0.661	0.474	0.687	0.464	0.693	0.461	0.655	0.475	0.631	0.482	0.606	0.489	0.654	0.476	0.602	0.49	0.575	0.494
feeling equal Catalan than Spanish, more Catalan or Catalan only	0.97	0.17	0.971	0.169	0.949	0.22	0.953	0.211	0.952	0.214	0.939	0.239	0.95	0.218	0.954	0.209	0.945	0.228
Catalonia independent State (0-1)	0.476	0.499	0.611	0.488	0.614	0.487	0.562	0.496	0.534	0.499	0.499	0.500	0.537	0.499	0.490	0.500	0.460	0.498
deficient autonomy of Catalonia (0-1)	0.779	0.415	0.805	0.396	0.809	0.393	0.772	0.419	0.775	0.417	0.761	0.427	0.768	0.422	0.758	0.428	0.726	0.446
binary independence preference (0-1)	--	--	--	--	0.626	0.484	0.635	0.481	0.624	0.484	0.601	0.490	0.642	0.479	0.599	0.490	0.569	0.495
Catalan as self-identification language (0-1)	0.749	0.434	0.747	0.434	0.715	0.452	0.670	0.470	0.581	0.493	0.563	0.496	0.571	0.495	0.561	0.496	0.548	0.498
years of compulsory education in Catalan (0-10)	3.272	4.222	3.310	4.260	3.556	4.248	3.841	4.323	4.131	4.368	4.042	4.341	4.317	4.410	4.234	4.382	4.467	4.401
both parents born in the rest of Spain (0-1)	0.217	0.413	0.219	0.414	0.213	0.410	0.229	0.421	0.271	0.445	0.313	0.464	0.278	0.448	0.305	0.460	0.299	0.458
native Spanish speaker (0-1)	0.318	0.466	0.302	0.459	0.281	0.450	0.334	0.472	0.405	0.491	0.454	0.498	0.410	0.492	0.467	0.499	0.470	0.499
female (0-1)	0.501	0.500	0.509	0.500	0.508	0.500	0.507	0.500	0.507	0.500	0.508	0.500	0.505	0.500	0.510	0.500	0.514	0.500
age (18-84)	46.53	17.52	46.99	17.34	47.26	16.68	46.78	16.65	46.36	16.57	47.38	16.52	47.13	16.57	47.89	16.37	47.92	16.40
TV3 for political information (0-1)	0.564	0.496	0.530	0.499	0.496	0.500	0.452	0.498	0.410	0.492	0.427	0.495	0.425	0.494	0.419	0.493	0.409	0.492

Table A3 of the Appendix displays the pairwise correlation between the different indicators of identity and political preferences that we use in this paper. All of them are positively correlated with each other (also statistically significant). It is worth-while noticing that the indicators of linguistic national identity exhibit relatively low correlation coefficients. In particular, Catalan as the language of self-identification reaches the highest correlation coefficient, 58%, with feeling “More Catalan or only Catalan”. Similarly, the correlation coefficients of political outcomes and identity measures are never higher than 75%, which is the value reached by the dummy for supporting independence in a binary choice the indicator “Catalonia as independent State” (d).<sup>20</sup> This

<sup>20</sup> Obviously, the 1-5 scale for the feeling of belonging is highly associated with the indicator for feeling only Catalan.

evidence indicates that, in general, indicators of identity reflect different facets of it, and political preferences are not only influenced by identity considerations.

As for control variables, around 25% of the pooled sample have both parents born in the rest Spain<sup>21</sup> and the share of individuals whose native language is Spanish (only) ranges between 28% and 47% (being this fraction higher in the last waves). Average age fluctuates around 46 and the sample is quite balanced in terms of gender. Finally, on average, 46% of individuals in the pooled sample watches TV3 (the main TV channel in Catalan, launched in 1983) to get informed about political issues.

In order to visualize the evolution of the outcomes across the cohorts, Figures 1-7 displays binned scatter plot of their averages by year of birth, together with a local linear smooth plot. We also allow for a discontinuity at 1970, corresponding to the first cohort of individuals exposed to Catalan during compulsory education. The evolution of the 1-5 scale reflecting the feeling of belonging shows a clear negative trend (indicating less attachment to the Catalan identity) for individuals born after 1940, which becomes stronger for the youngest cohorts. However, there is no any clear change in this tendency around the pivotal year. The evidence is different when we plot the fraction of individuals answering they feel only Catalan: the proportion is remains stable for individuals born between 1920 and 1970, but declines afterwards. The evolution of the fraction of individuals who feel more Catalan than Spanish or only Catalan, as well as the trend for the indicator that also includes those who consider themselves equally Catalan than Spanish follows the same decreasing pattern observed for the 1-5 scale of feeling of belonging. Again, there is no clear discontinuity for cohorts affected by the introduction of Catalan at school.

We instead obtained a reverted picture from the graph displaying the share of individuals who declare that Catalonia should be an independent State. Overall, there is a positive tendency across the cohorts. However, this is mostly driven by the oldest (until 1940) and youngest cohorts (since 1990, which also display a higher degree of dispersion). Nevertheless, the evolution of this outcome is quite stable for intermediate cohorts, including those who were initially affected by the LNA reform of 1983. The tendency observed for the variable regarding the level of autonomy that Catalonia should have with respect to the Central State is also negative for individuals born after 1950 and a similar picture is observed for the fraction of independence supporters. For both variables, there is no change in the trend across the cohorts from 1970. Finally, the share of individuals who declare that Catalan (only) represents their self-identification language shows a pronounced negative tendency across the cohorts. However, this negative trend appears to be attenuated for the first cohorts affected by the LNA reform, although it is reinforced for individuals born after 1990.

---

<sup>21</sup> This variable takes the value one also when one parent was born in other Spanish regions but the other was born in other countries. On the contrary, it is set to missing when both parents were born outside Spain (410 observations among individuals who satisfy the previous criteria).

This descriptive evidence suggests that, if any, the implementation of the LNA reform only moderated the “extreme” feeling of Catalan identity, but also the decline in the tendency across birth cohorts of adopting Catalan as self-identification language, without exerting any clear effect on the other measures of national identity and political preferences. However, one may claim that these results could be confounded by several factors such as age differences, changes over time and other potential spurious cohort trends. In the next section, we present the empirical strategy that we follow to estimate the causal effect of the exposure to Catalan at school induced by the implementation of the LNA reform.

#### 4 Empirical Strategy

In order to estimate the causal effect induced by the LNA reform of 1983 on identity formation, we exploit variation in potential exposure by birth cohort following the same approach proposed by CF-M.<sup>22</sup> More specifically, with the aim of relying on an exogenous measure of exposure to Catalan at school, we construct a variable capturing the years of exposure to the language-in-education policy during compulsory education. In Spain, primary education starts at age 6. For cohorts of individuals born until 1982, the duration of compulsory education, which comprises primary and lower secondary education, was of 8 years (i.e. until age 14). This was the framework established by the General Law of Education (LGE) of 1970. Therefore, individuals born before 1970 were never exposed to Catalan during compulsory schooling. On the other hand, the LNA reform was sharply introduced across all grades in primary and lower secondary education in 1983, implying that for cohorts already attending school in that academic year the actual amount of exposure depends on the remaining number of years before completion of compulsory education. This implies that those born between 1971 and 1976 received part of compulsory education only in Spanish, and part in Catalan and Spanish, generating partial exposure to the reform. Moreover, individuals born since 1977 were entirely schooled under the new bilingual system and received 8 years of compulsory education in Catalan (and Spanish). However, a subsequent reform that took place in 1990 extended compulsory education from 8 to 10 years (Law of General Ordering of the Educational System, LOGSE), translating into longer exposure to bilingual schooling for cohorts born in 1983 and onwards.

Therefore, to capture this differential intensity of exposure to the reform, we construct the number of (potential) years of exposure to Catalan during compulsory education as a stepwise function of the birth year of individual  $i$  ( $\tau(i)$ ):

---

<sup>22</sup> Notice that a variant of this approach has been used by Caminal and Di Paolo (2019) and by Cappellari et al. (2021) in order to estimate the causal effect of oral fluency in Catalan among native Spanish speakers on partnership formation and intergenerational language transmission, respectively.

$$E_{\tau(i)} = \begin{cases} 10 & \text{if } \tau(i) \geq 1983 \\ 8 & \text{if } 1977 \leq \tau(i) < 1983 \\ \tau(i) - 1969 & \text{if } 1970 \leq \tau(i) < 1977 \\ 0 & \text{if } \tau(i) < 1970 \end{cases} \quad (1)$$

This represents the variable of interest, that is used as main regressor of the following equation to be estimated by OLS:

$$Y_i = \alpha + \beta E_{\tau(i)} + f(\text{age}_i) + \delta' X_i + \theta_{yw} + \varepsilon_i. \quad (2)$$

Here  $Y_i$  represents each of the different identity measures,  $f(\text{age}_i)$  is a flexible polynomial of age at the time of the survey (to control for age-differences in the outcomes),  $X_i$  is a vector of exogenous controls at the individual level (initially only gender) and the term  $\theta_{yw}$  contains year-wave specific fixed effects that account for potential trends in the outcomes over time. The coefficient of interest is  $\beta$ , which captures the impact of each additional year of compulsory exposure to Spanish-Catalan bilingualisms at school. Because the  $E_{\tau(i)}$  variable is defined according to birth cohort, we cluster the standard errors at the year of birth level. Notice also that the availability of repeated cross-sections of data enable us to properly control for age effects in equation (2), which are separately identified with respect to the variable of interest. This is because the sample contains individuals born in the same year who have different aged when they participated in the survey. On the contrary, the regressions reported in CF-M are based on a single cross-section, implying that the coefficients of the age polynomial are separately identified from the exposure coefficient only on the basis of strong functional form assumptions.

Although the measure of exposure to Catalan during compulsory education  $E_{\tau(i)}$  is clearly exogenous, as it depends exclusively on year of birth and on the duration of compulsory schooling, it might not be a variable to capture the true causal effect of the LNA reform on national identity. This is because  $E_{\tau(i)}$  may capture changes across the cohorts that might affect identity formation through other channels than language exposure during compulsory education, generating spurious estimates. Therefore, we apply two different types of falsification tests based on placebo reforms for groups of individuals who were never really exposed to Catalan at school.

First, similarly to what was done by CF-M, we consider individuals who satisfy the restriction to form part of the main sample (i.e. born in Catalonia or in the rest of Spain, but migrated at age 6 or earlier) but excluding those who were partially or entirely exposed to the LNA reform (i.e. individuals born from 1970). Therefore, using these sample of “older” cohorts of individuals, we specify a battery of measures of placebo exposure to the fake reform that were hypothetically implemented “ $l$ ” years before the true reform of 1983. This approach would reflect the presence of pre-existing spurious cohort trends in the outcomes.

Second, as mentioned in the previous section, we retained individuals born in other Spanish regions or outside Spain born in the same cohorts than our baseline sample, but who migrated to Catalonia after age 16 and, therefore, were never exposed to Catalan at school. Using this placebo sample, we estimate the same regression described in equation (1) treating these individuals “as if” they were schooled in Catalonia. This additional and novel placebo estimation enables checking whether the main results based on the baseline sample are affected by contemporaneous spurious trends across the cohorts.

Finally, we also try to consider that the LNA reform not only affected the language of instruction, but also other aspect related to the recovery of the use of Catalan in the society. Specifically, we are mostly concerned about the introduction of a public TV channel in Catalan (*Televisió de Catalunya, TV3*), which was launched exactly in the same year than the introduction of Catalan as medium of instruction (1983). Our concern, also noticed by CF-M, comes from the fact that exposure to this new TV channel could have affected the formation of political identity independently of the language-in-education policy component of the LNA reform. Therefore, we exploit the information about the TV channel that individuals are used to watch to keep informed about political issues. More concretely, we constructed a dummy that is equal to one if the individuals declare they are used to follow political information through TV3 and zero otherwise and use this indicator as alternative outcome for equation (1).<sup>23</sup>

## 5 Results

The estimates reported in Table 2 refer to the standard indicator of national identity, the feeling of belonging, also used by CF-M. In the first column we show the results obtained from an OLS regression with the entire 1-5 scale of feeling of belonging as dependent variable. Our results are clearly at odds with the evidence reported by CF-M, who considered exactly the same dependent variable. The estimate that we obtain for the exposure variable is virtually zero and not statistically significant. The last column corresponds to the binary variable “Only Catalan, or More or Equal Catalan” as the dependent variable, also used by CF-M. The results are qualitatively similar to theirs (but smaller in size), indicating that each additional year of exposure increases the propensity to feel “at least” as Catalan as Spanish. In our view, this variable represents a blurred indicator of national identity, since it combines clear Catalan-oriented options with the neutral alternative. Moreover, such a variable is equal to one for 95% of the estimation sample (87% in CF-M using data from 2001). Indeed, if the identity dummy is constructed in different ways, considering only clear-cut options, the results are substantially different. First, the exposure coefficient is equal to zero when the dummy captures individuals who feel more Catalan than Spanish or only Catalan. Second, the coefficient becomes negative, and statistically significant, when the identity indicator takes the value one only for individuals who feel exclusively Catalan. This is indeed consistent with the

---

<sup>23</sup> This contrasts with the approach followed by CF-M, who used the same variable as additional control in their regressions.



graphical evidence reported in Figure 2, suggesting that the LNA reform moderated the intensity of national identity feelings, at least according to this specific variable.

**Table 2: compulsory education in Catalan and feeling of belonging**

Variable: feeling of belonging	1-5 scale	Catalan only	more Catalan or only Catalan	equal Catalan than Spanish, more Catalan, only Catalan
years of compulsory education in Catalan	0.001 (0.002)	-0.003*** (0.001)	0.002 (0.001)	0.002*** (0.001)
mean of the dependent variable	3.835	0.292	0.608	0.954
adjusted R-squared	0.022	0.024	0.017	0.006
number of observations	77678	77678	77678	77678

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

The baseline results concerning the other measures of identity, as well as political preferences, that we consider in this paper are shown in Table 3. The first three columns correspond to three indicators of political preferences, and the last one to language identity. The first column contains the results for the variable regarding preferences for the organization of the state and the derived indicator that takes the value one if the individual consider that Catalonia should be an independent state rather than belonging to Spain as a region, an autonomous community or a federal state. Each year of exposure to Catalan during compulsory schooling reduces the propensity to claim that Catalonia should be an independent state. However, when we consider as outcomes the indicator for considering that Catalonia has a deficient level of autonomy the coefficient of exposure to the LNA reform reverts to positive, as happens when the dependent variable consists in the dummy for supporting independence in a binary choice.

**Table 3: compulsory education in Catalan and other identity measures**

Variable:	Catalonia independent State	deficient autonomy of Catalonia	binary independence of preference	Catalan as self- ident. language
years of compulsory education in Catalan	-0.004*** (0.001)	0.003** (0.001)	0.004** (0.002)	0.007*** (0.002)
mean of the dependent variable	0.383	0.740	0.613	0.648
adjusted R-squared	0.087	0.027	0.009	0.081
number of observations	74951	75481	22716	36112

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

With respect to linguistic identity, exposure to the education reform has fostered the adoption of Catalan as a language of self-identification. Thus, in line with the graphical evidence, the implementation of the LNA reform mitigated the negative trend in the propensity to adopt (only) Catalan as self-identification language.

Summarizing, placing Catalan as the main language of instruction in compulsory education has stimulated the adoption of Catalan as the language of self-identification. However, other indicators of national identity have not been affected in the same direction. In particular, our favorite measure (that includes the options Only Catalan, or More Catalan than Spanish) shows no effect. Moreover, if we focus on the most biased feeling (Only Catalan) the effect of the reform is negative. As far as the political preferences are concerned the results are mixed. In particular, the sign of the effect of the reform on preferences towards independence seem to depend on the number of alternatives. Thus, so far we have found no clear evidence in support of the hypothesis that the education reform has promoted the Catalan identity and enhanced the popular support for the political goals of the pro-Catalan movement.

### **5.1 Robustness checks**

In order to validate our results, we present two different falsification exercises aimed at showing that the exposure variable is not capturing spurious cohort trends. Focusing on the indicator for feeling exclusively Catalan, Table 4 shows the results obtained by specifying placebo exposure variables defined “as if” the (fake) reform was implemented  $X$  years before the real one of 1983. We therefore run the same regression than equation 2, considering as estimation sample observations from individuals who satisfy all the conditions for forming part of the main sample, but were born before 1970. In this way, we retain only individuals who were never exposed to Catalan during compulsory education. In the first column of the first table, we also report the baseline results for comparison. As it can be appreciated, the coefficients of the fake exposure variable are all virtually zero, both in terms of size and statistical significance, for any of the fake reforms (from 13 to 20 years before the real one). This results points out that the negative coefficient of exposure to the LNA reform in the regression for the indicator or feeling Catalan only is not driven by pre-existing cohorts trends unrelated to the policy change.

**Table 4: falsification using older cohorts and fake reforms – feeling only Catalan**

	baseline	1970	1969	1968	1967	fake reform in			
						1966	1965	1964	1963
years of compulsory education in Catalan	-0.003*** (0.001)								
fake years of compulsory education in Catalan		0.000 (0.002)							
fake years of compulsory education in Catalan			0.000 (0.002)						
fake years of compulsory education in Catalan				-0.000 (0.002)					
fake years of compulsory education in Catalan					-0.001 (0.002)				
fake years of compulsory education in Catalan						-0.002 (0.002)			
fake years of compulsory education in Catalan							-0.002 (0.002)		
fake years of compulsory education in Catalan								-0.002 (0.002)	
fake years of compulsory education in Catalan									-0.001 (0.002)
adjusted R-squared	0.024	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
number of observations	77678	43661	43661	43661	43661	43661	43661	43661	43661

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

As second falsification exercise, we instead use individuals born in the same cohorts than those forming part of our main estimation sample, but were born outside Catalonia (either in the rest of Spain or in other country) and migrated to Catalonia being older than 16. Therefore, they were never exposed to Catalan during compulsory education. Using this alternative placebo sample, we estimate equation 2, but this time the effect of the variable of interest would capture any possible contemporaneous spurious trend across the cohorts. Table 5 shows the results for all the identity feeling outcomes and, reassuringly, for none of them we obtain estimates that are somehow different than zero both in terms of size and statistical significance, which speaks in favor of the validity of the causal interpretation of our results.

**Table 5: Falsification using internal and foreign migrants**

Variable: feeling of belonging	1-5 scale	Catalan only	more Catalan or only Catalan	equal Catalan than Spanish, more Catalan, only Catalan
years of compulsory education in Catalan	-0.001 (0.006)	-0.001 (0.001)	0.002 (0.002)	0.002 (0.003)
mean of the dependent variable	2.643	0.036	0.135	0.676
adjusted R-squared	0.010	0.005	0.008	0.012
number of observations	12278	12278	12278	12278

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

Using the baseline sample and the last placebo sample including internal and foreign migrants, in Table 6 we also show that exposure to Catalan at school is significantly negatively related to the propensity to watch the news about political issues in TV3, the Catalan TV launched in the same year than the introduction of Catalan as medium of instruction. This negative sign is against the possible interpretation that the exposure coefficient of the main regression is driven by the contemporaneous availability of a TV channel in Catalan.

**Table 6: regression for Catalan TV - main sample and placebo sample with internal and foreign migrants**

	main sample	placebo sample
years of compulsory education in Catalan	-0.003** (0.001)	-0.005** (0.002)
mean of the dependent variable	0.518	0.176
adjusted R-squared	0.038	0.030
number of observations	77678	12278

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender and year-wave fixed effects.

Finally, because the validity of our identification strategy strongly relies on the correct specification of the age polynomial to control for age differences in the outcome, which may confound the effect of exposure to the language-in-education reform, we show the results obtained under different specifications of the age polynomial (ranging from 1 to 4). The results are reported in Table 7 and show that the main evidence regarding the outcome “feeling only Catalan” is generally unaffected by the order of the age polynomial, which is a further evidence regarding the stability of our main evidence.

**Table 7: robustness to the order of the age polynomial - feeling only Catalan**

	Order of the age polynomial			
	(1)	(2)	(3)	(4)
years of compulsory education in Catalan	-0.002** (0.001)	-0.003** (0.001)	-0.003*** (0.001)	-0.004*** (0.001)
age	0.001*** (0.000)	-0.000 (0.001)	0.000 (0.003)	0.016 (0.011)
age <sup>2</sup>		0.000 (0.000)	-0.000 (0.000)	-0.001 (0.000)
age <sup>3</sup>			0.000 (0.000)	0.000 (0.000)
age <sup>2</sup>				-0.000 (0.000)
adjusted R-squared	0.024	0.024	0.024	0.024
number of observations	77678	77678	77678	77678

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender and year-wave fixed effects.

## 5.2 Heterogeneous analysis

In this section, we show the results of the analysis of heterogeneous effects of the LNA reform on identity formation. The first dimension that we consider is the regional origins of the parents. Specifically, we analyze whether exposure to Catalan during compulsory education had a differential effect on the different measures of identity according to whether the parents were both born in other Spanish regions, or at least one of them was born in Catalonia. In order to do that, we split the sample according to this binary indicator for parental origins and repeat the baseline estimation (equation (2)).

The results regarding the variable capturing identity feelings, displayed in Table 8, indicate that the effect of exposure to Catalan at school is strongly heterogeneous according to parent's origins. Referring to the 1-5 scale, the effect of exposure is positive (i.e. more attachment to the Catalan identity) for individuals with both parents born in other Spanish regions, but negative for those with both parents born in Catalonia. Among the latter, exposure to Catalan at school also reduces the "extreme" feeling of Catalan identity, as well as the propensity to feel "only Catalan" or "more Catalan than Spanish". On the other hand, the exposure variable has a positive coefficient on the probability to feel at least as Catalan than Spanish for individuals with both parents born outside Catalonia.

**Table 8: heterogeneous effects by parents' origins - feeling of belonging**

Variable: feeling of belonging	1-5 scale		Catalan only		more Catalan or only Catalan		equal Catalan than Spanish, more Catalan, only Catalan	
	both born outside Catalonia	at least one born in Catalonia	both born outside Catalonia	at least one born in Catalonia	both born outside Catalonia	at least one born in Catalonia	both born outside Catalonia	at least one born in Catalonia
years of compulsory education in Catalan	0.007** (0.003)	-0.017*** (0.004)	-0.001 (0.002)	-0.009*** (0.002)	0.002 (0.002)	-0.007*** (0.002)	0.004*** (0.001)	-0.000 (0.001)
adjusted R-squared	0.023	0.030	0.023	0.032	0.023	0.019	0.008	0.006
number of observations	19240	57922	19240	57922	19240	57922	19240	57922

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

As for the other outcomes (Table 9), the results regarding the dummy for declaring that Catalonia should be an independent state indicate that the negative effect of the LNA reform is significant and negative only among individuals with at least one parent born in Catalonia. Moreover, the LNA reform increased the likelihood to adopt Catalan as self-identification language among individuals with both parents born outside Catalonia become, while the opposite effect is observed for those with at least one parent born in Catalonia. The results for other outcomes are inconclusive.

**Table 9: heterogeneous effects by parents' origins - other identity measures**

Variable:	Catalonia independent State		deficient autonomy of Catalonia		binary independence preference		Catalan as self-ident. language	
	both born outside Catalonia	at least one born in Catalonia	both born outside Catalonia	at least one born in Catalonia	both born outside Catalonia	at least one born in Catalonia	both born outside Catalonia	at least one born in Catalonia
years of compulsory education in Catalan	-0.002 (0.002)	-0.008*** (0.002)	0.003 (0.002)	-0.002 (0.001)	-0.005 (0.004)	-0.001 (0.002)	0.006** (0.003)	-0.005*** (0.002)
adjusted R-squared	0.046	0.114	0.028	0.030	0.005	0.019	0.074	0.095
number of observations	18175	54939	18184	55363	6058	16475	8889	26012

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

Table 10 shows the regressions for the different definitions of identity feeling with heterogeneous effects by native language, again by estimating the model separately for the subsample of Catalan and Spanish (or mixed) native speakers. The results indicate that exposure to Catalan at school has a negative coefficient on the 1-5 scale of national identity for native Catalan speakers, who also become less likely to feel “only Catalan” or at least “more Catalan than Spanish”.

**Table 10: heterogeneous effects by native language - feeling of belonging**

Variable: feeling of belonging	1-5 scale		Catalan only		more Catalan or only Catalan		equal Catalan than Spanish, more Catalan, only Catalan	
	Spanish or Spanish & Catalan	Catalan	Spanish or Spanish & Catalan	Catalan	Spanish or Spanish & Catalan	Catalan	Spanish or Spanish & Catalan	Catalan
years of compulsory education in Catalan	0.007 (0.005)	-0.014*** (0.003)	0.000 (0.002)	-0.013*** (0.002)	0.004 (0.003)	-0.003* (0.002)	0.002 (0.002)	0.001 (0.001)
adjusted R-squared	0.022	0.018	0.017	0.020	0.019	0.009	0.013	0.002
number of observations	13043	21878	13043	21878	13043	21878	13043	21878

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

Table 11 displays the results for other outcomes, with separate estimations according to native language. These are quite similar to those obtained after splitting the sample according to parental origins. Specifically, native Catalan speakers exposed to the LNA reform are less likely to claim that Catalonia should be an independent state, while no effect is found among native Spanish speakers. Moreover, the former are less likely to declare that Catalan is their self-identification language after being exposed to bilingualism at school, while the opposite effect emerges among native Spanish speakers, who are more likely to become language switchers.

**Table 11: heterogeneous effects by native language – other identity measures**

Variable:	Catalonia independent State		deficient autonomy of Catalonia		binary independence preference		Catalan as self-ident. language	
Native language:	Spanish or Spanish & Catalan	Catalan	Spanish or Spanish & Catalan	Catalan	Spanish or Spanish & Catalan	Catalan	Spanish or Spanish & Catalan	Catalan
years of compulsory education in Catalan	0.001 (0.002)	-0.007*** (0.002)	0.001 (0.002)	-0.001 (0.001)	-0.000 (0.003)	-0.000 (0.002)	0.004** (0.002)	-0.003* (0.001)
adjusted R-squared	0.012	0.025	0.009	0.011	0.009	0.007	0.071	0.035
number of observations	12257	20818	12243	21032	8953	13550	13014	21849

Note: Standard errors clustered at the year of birth level; \*\*\* significant at 1%, \*\* significant at 5%, \* significant at 10%. All regressions include controls for gender, a third-order age polynomial and year-wave fixed effects.

## 6 Conclusions (preliminary)

In this paper we have examined the effect of a language-in-education reform that took place in Catalonia, starting in 1983, and that placed the Catalan language at centerstage of the school system. We have found no solid evidence in support of the hypothesis that such reform has interfered with the identity formation of its pupils and altered the subsequent political preferences to make them more favorable to Catalan-oriented goals.

More specifically, if we focus on our favorite indicators of Catalan identity ((i) Feeling Catalan only, and (ii) More Catalan than Spanish or Catalan only), the exposure to the reform has actually had a negative effect on promoting the Catalan identity, especially among native Catalan speakers, or those subjects with Catalan origins. If anything, the reform has moderated the extreme feelings and promoted the dual identity (as Catalan as Spanish). Our indicator of linguistic identity (Catalan as the language of self-identification) provides a somewhat different picture. The exposure to the reform has had a positive effect, at least for subjects with non-Catalan family origins. In contrast, the effect on native Catalan speakers and, more especially, on native Spanish speakers with Catalan origins are negative. It is remarkable that the two dimensions of identity (national and linguistic) are only moderately correlated. Therefore, embracing Catalan as the language of self-identification does not imply being assimilated into a social group defined in national or political terms. In other words, even though we often consider the language as the marker of ethnic or national identity, such a link is more complex than initially expected. Such evidence is in line with the one reported by anthropologists and sociolinguists, according to which Catalan is less and less perceived as the language of and ethnic group and more and more as “the language of everyone” (see Woolard and Gahng, 1990 and Newman and Trenchs-Parera, 2015).

As far as our indicators of political preferences are concerned, the results offer a similar picture: the effects of the exposure are either zero (statistically not significant) or negative.

It is perhaps surprising that native Catalan speakers is the group with the strongest negative effect of exposure on their support for independence.

We can only speculate about the reasons behind such null and even negative effects. On the one hand, it could be the case that policy makers did not intend or did not have access to the necessary means to set up a nationally-biased environment. This reform took place under circumstances that had little to do with the nation-building efforts of many states in the XIX and XX centuries. On the other hand, even if such a biased framework was present, and the school did try to interfere with the identity formation of individuals, such attempt could have put in motion opposite efforts by other institutions, including the family. Such backlash would be analogous to the evidence reported by Fouka (2020).

## References

Alba, R., J. Logan, A. Lutz, B. Stults (2002). "Only English by the Third Generation? Loss and Preservation of the Mother Tongue among the Grandchildren of Contemporary Immigrants". *Demography*, 39 (3), 467-484.

Alesina, A. P. Giuliano, and B. Reich (2021). "Nation-building and Education". *The Economic Journal* 131 (August), 2273-2303.

Alesina, A., B. Reich, and A. Riboni (2020). "Nation-building, Nationalism, and Wars". *Journal of Economic Growth* 25 (4), 381-430.

Alesina, A., E. Spolaore, and R. Wacziarg (2000). "Economic Integration and Political Desintegration". *American Economic Review* 90 (5), 1276-96.

Aspachs-Bracons, O., I. Clots-Figueras, and P. Masella (2008). "Compulsory Language, Educational Policies, and Identity Formation". *Journal of the European Economic Association, Papers and Proceedings*, vol. 6, 433-444.

Bagues, M., & Roth, C. (2022). Interregional Contact and the Formation of a Shared Identity. *American Economic Journal: Economic Policy* (forthcoming).

Bandiera, O., M. Mohnen, I. Rasul, and M. Viarengo (2018). "Nation-Building through Compulsory Schooling during the Age of Mass Migration". *Economic Journal* 129 (167), 69-102.

Bazzi, S., A. Gaduh, A. Rothenberg, and M. Wong (2019). "Unity in Diversity? How Intergroup Contact can Foster Nation Building". *American Economic Review* 109 (11), 3987-4025.



- Cáceres-Delpiano, J., A. De Moragas, G. Facchini, and I. Gonzalez (2021). "Intergroup contact and nation building: Evidence from military service in Spain", *Journal of Public Economics* 201, 104477.
- Caminal, R., & Di Paolo, A. (2019). Your language or mine? The noncommunicative benefits of language skills. *Economic Inquiry*, 57(1), 726-750.
- Caminal, R., L. Cappellari, A. Di Paolo (2021). "Language-in-education, language skills and the intergenerational transmission of language in a bilingual society", *Labour Economics*, vol. 70, num. 101975.
- Cappellari, L., A. Di Paolo (2018). "Bilingual schooling and earnings: Evidence from a language-in-education reform". *Economics of Education Review*, 64, 90-101.
- Casas, A., F. Curci, and A. De Moragas (2022). "Checks and Balances and Nation Building: The Spanish Constitutional Court and Catalonia". CUNEF University.
- Clots-Figueras, I., and P. Masella (2013). "Education, Language and Identity". *The Economic Journal* 123 (August), F332-F353.
- Dehdari, S. and K. Gehring (2022). "The Origins of Common Identity: Evidence from Alsace-Lorraine". *American Economic Journal: Applied Economics* 14 (1), 261-292.
- Dougan (2007) Duggan, C. (2007). *The Force of Destiny: A History of Italy since 1796*. London: Penguin.
- Easterly, W. and R. Levine (1997). "Africa's Growth Tragedy: Policies and Ethnic Divisions". *The Quarterly Journal of Economics* 4 (November), 112, 1203-1250.
- Fouka, V. (2020). "Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I". *Review of Economic Studies* 87, 204-239.
- Ginsburgh, V. and S. Weber (2020). "The Economics of Language". *Journal of Economic Literature* 58 (2), 348-404.
- Hauk, E. and J. Ortega (2019). "Schooling, nation building and industrialization". *Journal of Theoretical Politics*, 33(1), 56-94.
- Hobsbawm, E. (1990). *Nations and Nationalism since 1780: Programme, Myth, Reality*. Cambridge UK: Cambridge University Press.
- Newman, M. and M. Trenchs-Parera (2015). Language Policies, Ideologies and Attitudes in Catalonia. Part 1: Reversing Language Shift in the Twentieth Century. *Language and Linguistics Compass* 9/7, 285-294.

Soysal, Y. N., & Strang, D. (1989). Construction of the first mass education systems in nineteenth-century Europe. *Sociology of education*, 277-288.

Ricento, T. (2013). "Language policy, ideology and attitudes in English-dominant countries. In R. Bayley, R. Cameron, and C. Lucas (Eds.), *The Oxford Handbook of Sociolinguistics*, 525-544. Oxford: Oxford University Press.

Ricento, T. (2003). "The Discursive Construction of Americanism". *Discourse and Society* 14 (5), 611-637.

Weber, E. (1979). *Peasants into Frenchmen: Modernization of Rural France, 1870-1914*. London: Chatto and Windus.

Woolard, K. and T. Gahng, (1990). Language Policies and Attitudes in Autonomous Catalonia. *Language in Society*, Vol. 19, No. 3 (Sep., 1990), pp. 311-330

Figure 1: feeling of belonging (1-5 scale, from "only Spanish" to "only Catalan")

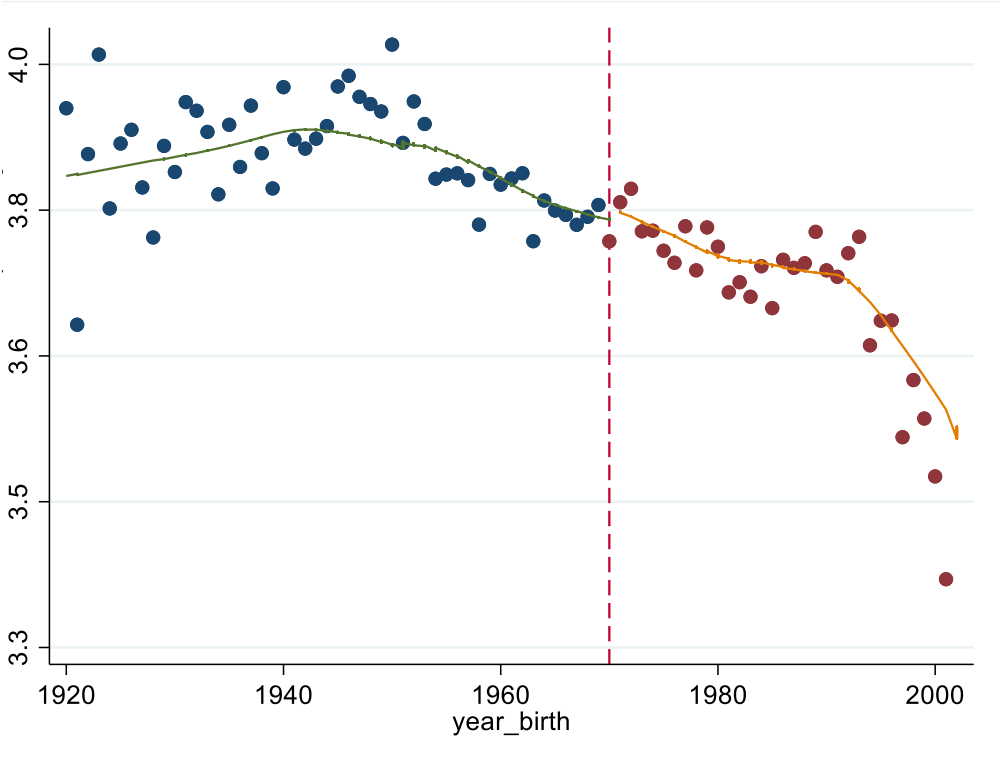
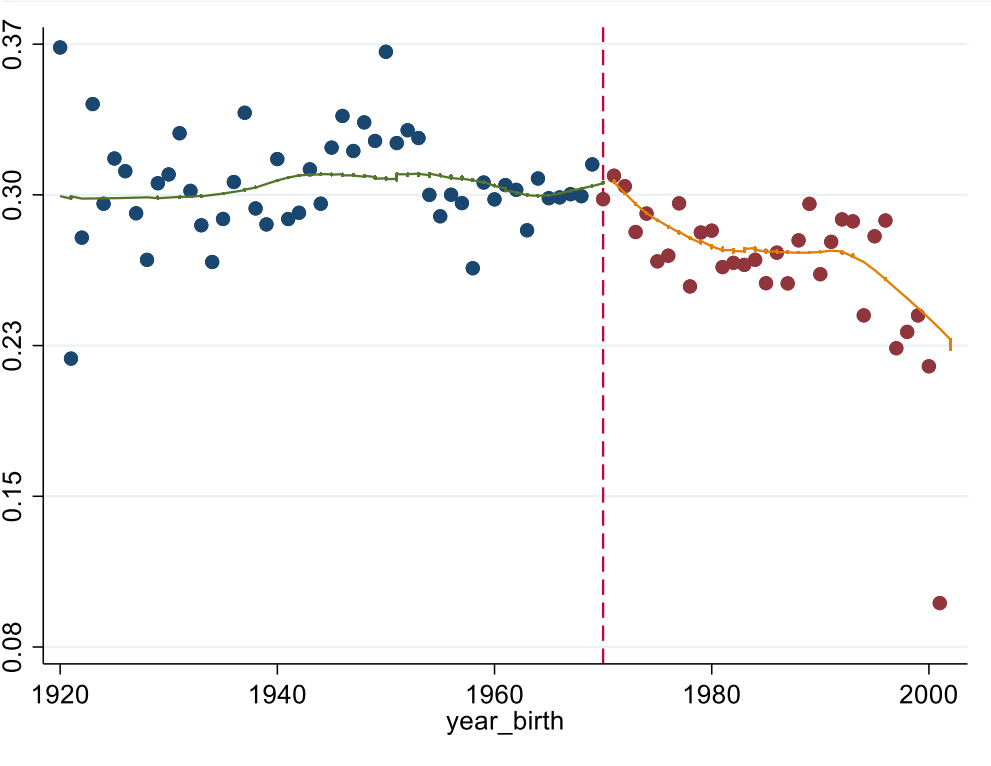
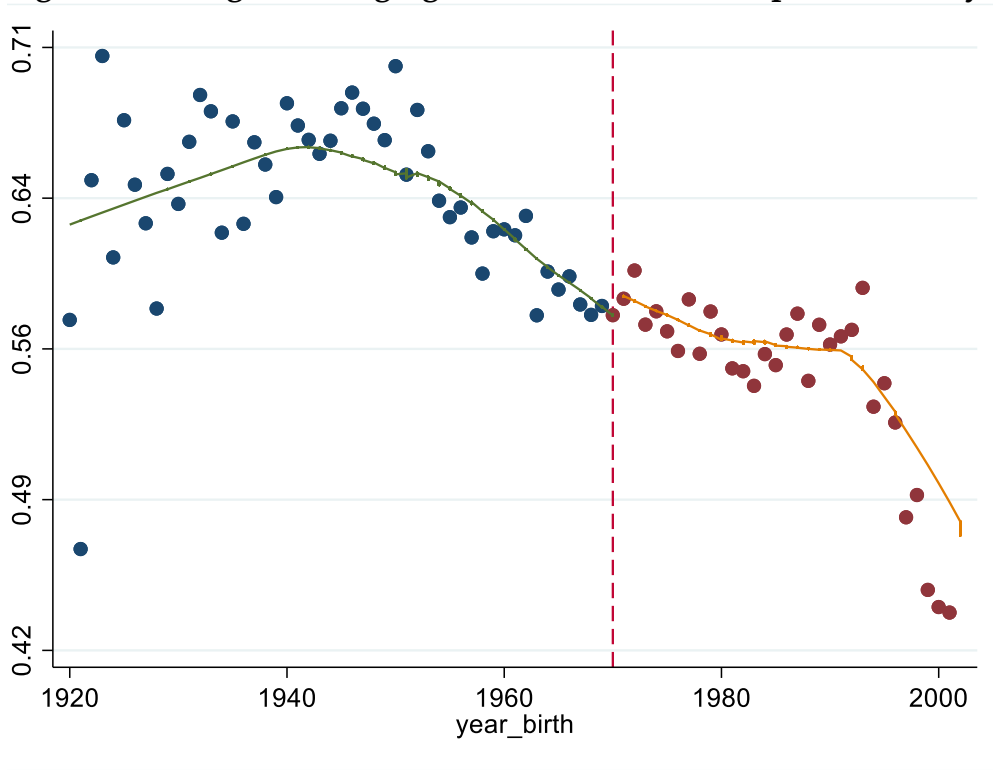


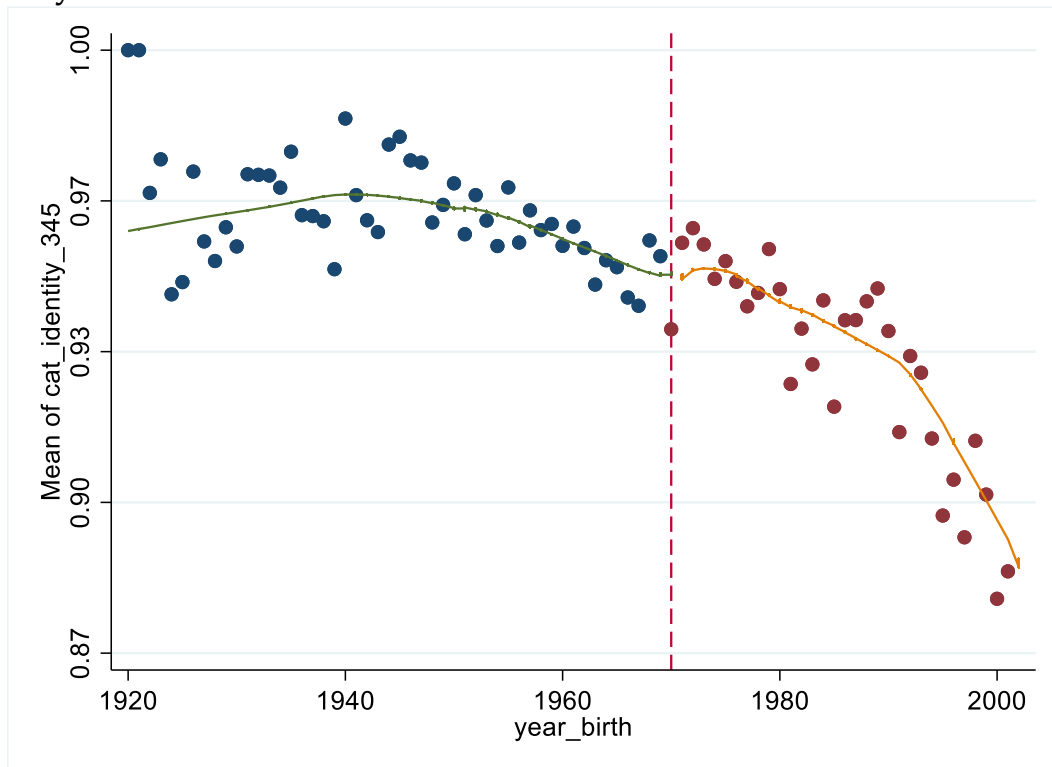
Figure 2: feeling of belonging = only Catalan



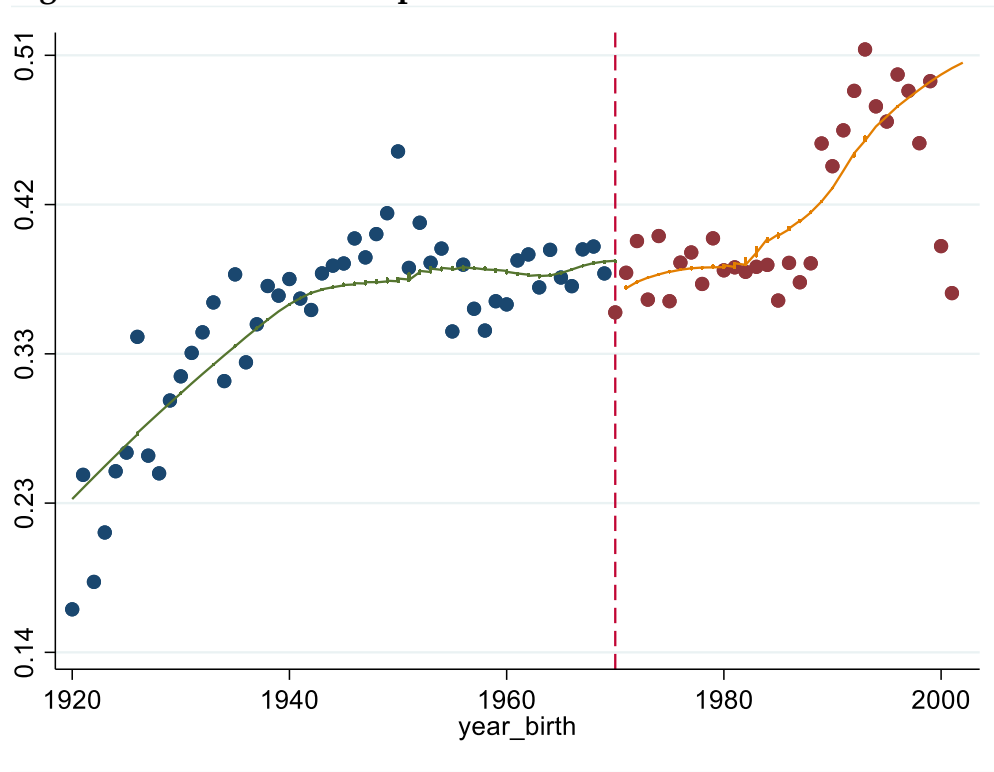
**Figure 3: feeling of belonging = more Catalan than Spanish or only Catalan**



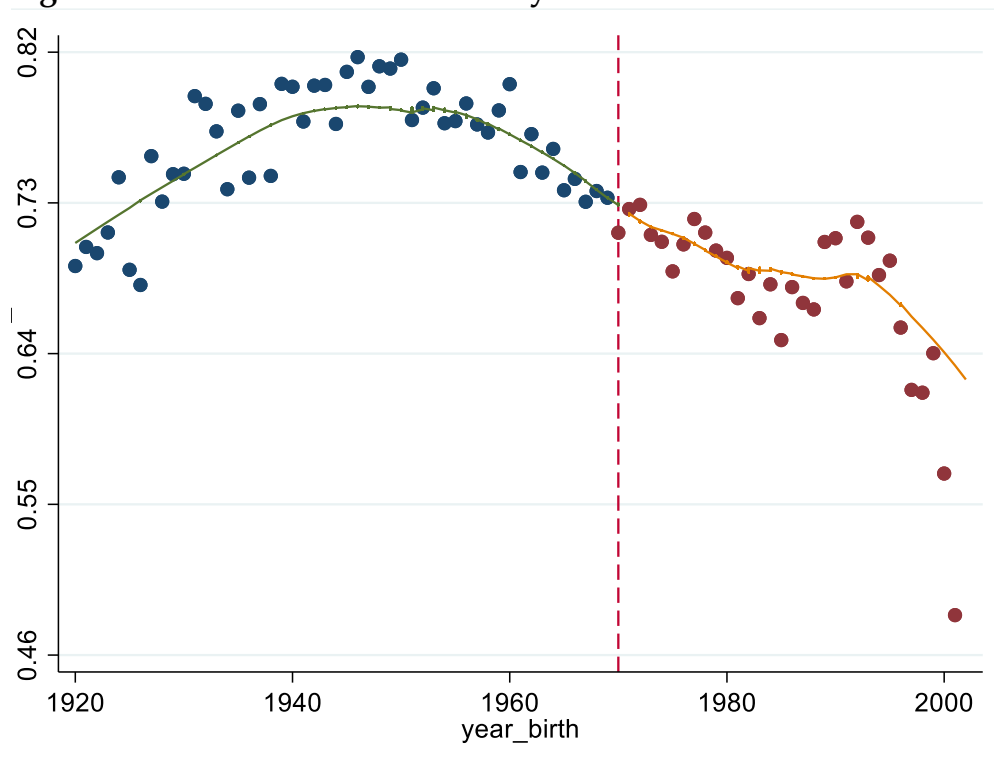
**Figure 4: feeling of belonging = as Catalan as Spanish, more Catalan than Spanish or only Catalan**



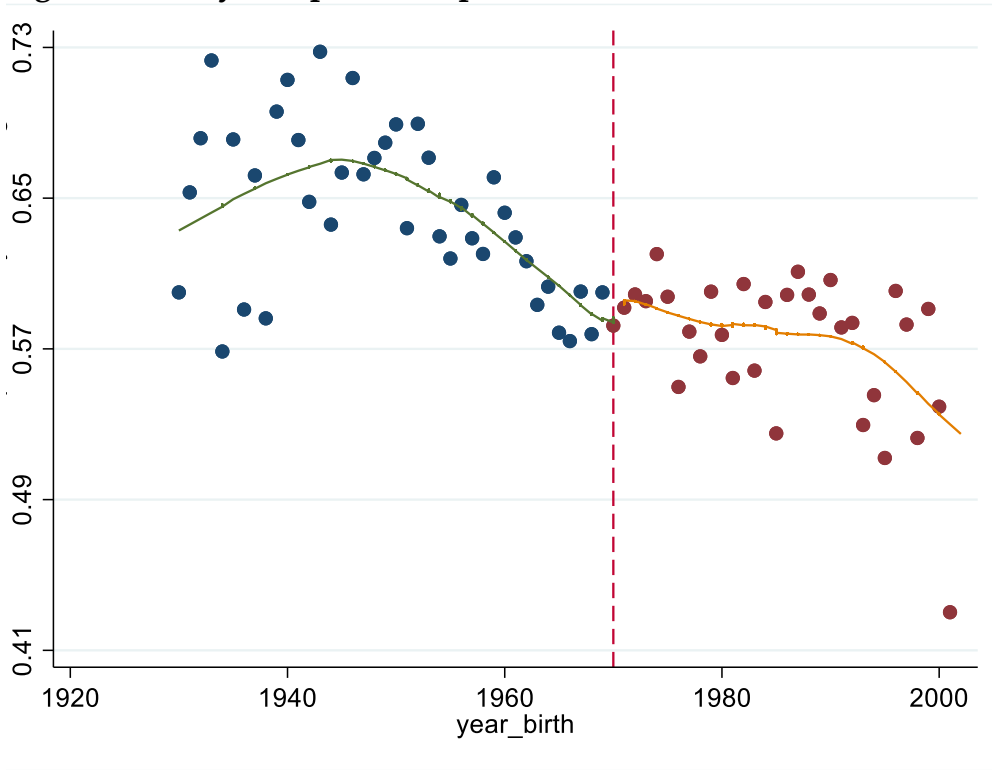
**Figure 4: Catalonia as independent State**



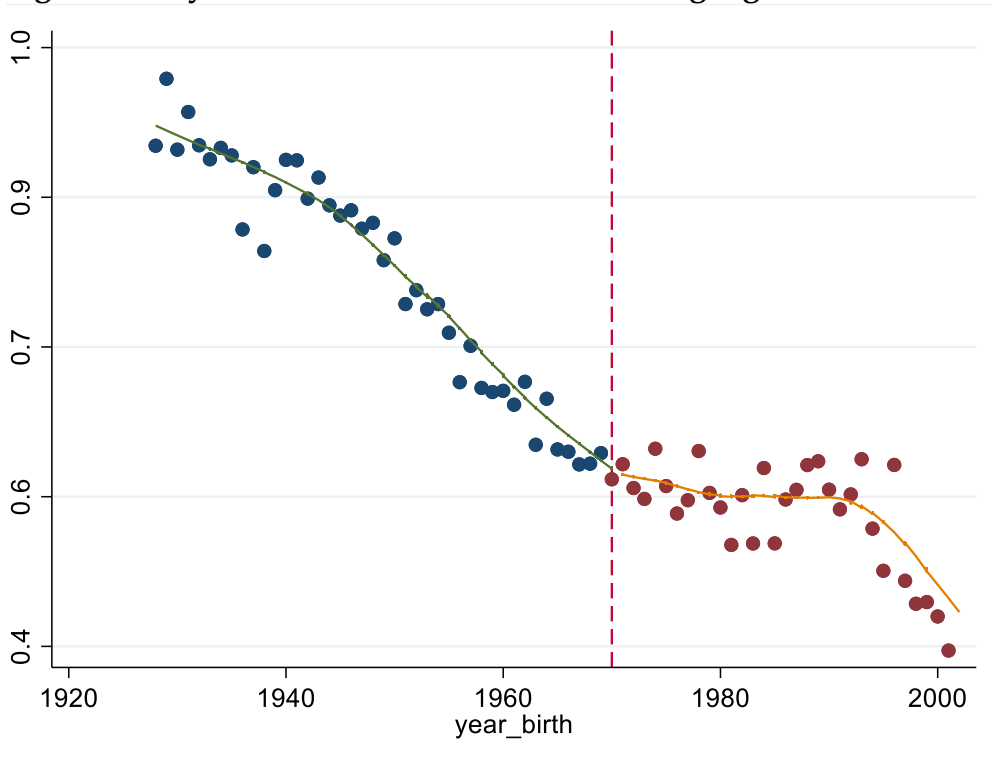
**Figure 5: deficient level of autonomy of Catalonia**



**Figure 6: binary independence preference indicator**



**Figure 7: only Catalan as self-identification language**



## Appendix

**Table A1: sample size by year and wave**

year/wave	1	2	3	4	Total
<b>2004</b>	1824	--	--	--	1824
<b>2005</b>	1900	1900	--	--	3800
<b>2006</b>	2100	2100	2100	2100	8400
<b>2007</b>	2200	2200	2200	2200	8800
<b>2008</b>	2000	2000	2000	2000	8000
<b>2009</b>	2000	2000	2000	2000	8000
<b>2010</b>	2000	2000	2000	2000	8000
<b>2011</b>	2500	2500	2500	--	7500
<b>2012</b>	2500	2500	2500	--	7500
<b>2013</b>	2000	2000	2000	--	6000
<b>2014</b>	2000	2000	--	--	4000
<b>2015</b>	2000	2000	2000	--	6000
<b>2016</b>	1500	1500	1500	--	4500
<b>2017</b>	1500	1500	1338	--	4338
<b>2018</b>	1500	1500	1500	--	4500
<b>2019</b>	1500	1500	1500	--	4500
<b>2020</b>	2000	2000	2000	--	6000
<b>Total</b>	33024	31200	27138	10300	101662

**Table A2: descriptive statistics for the placebo sample**

Variable	Mean	S.D.
feeling of belonging (1-5)	3.835	0.942
feeling Catalan only (0-1)	0.292	0.454
feeling more Catalan than Spanish or Catalan only	0.608	0.488
feeling equal Catalan than Spanish, more Catalan or	0.954	0.209
Catalonia independent State (0-1)	0.383	0.486
deficient autonomy of Catalonia (0-1)	0.740	0.439
binary independence preference (0-1)	0.613	0.487
Catalan as self-identification language (0-1)	0.648	0.478
years of compulsory education in Catalan (0-10)	3.288	4.172
both parents born in the rest of Spain (0-1)	0.249	0.433
native Spanish speaker (0-1)	0.374	0.484
female (0-1)	0.507	0.500
age (18-84)	46.36	17.17
TV3 for political information (0-1)	0.518	0.500

**Table A3: identity indicators' correlation matrix**

	1	2	3	4	5	6	7	8
1 feeling of belonging (1-5)	1							
2 feeling Catalan only	0.7933	1						
3 feeling more Catalan or Catalan only	0.8519	0.5154	1					
4 feeling equal Catalan than Spanish, more Catalan or Catalan only	0.5219	0.1402	0.272	1				
5 Catalonia independent State	0.5634	0.5421	0.4966	0.1434	1			
6 deficient autonomy of Catalonia	0.4414	0.2942	0.4473	0.214	0.3217	1		
7 binary independence preference	0.7129	0.5581	0.7433	0.2655	0.828	0.5051	1	
8 Catalan as self-identification language	0.5751	0.4327	0.576	0.2446	0.4776	0.347	0.5403	1